January 1991

Surface Structure, Intonation, and Meaning in Spoken Language

Mark Steedman
University of Pennsylvania

University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-91-12.

Surface Structure, Intonation, and Meaning in Spoken Language

Abstract

The paper briefly reviews a theory of intonational prosody and its relation syntax, and to certain oppositions of discourse meaning that have variously been called "topic and comment", "theme and rheme", "given and new", or "presupposition and focus". The theory, which is based on Combinatory Categorial Grammar, is presented in full elsewhere. the present paper examines its consequences for the automatic synthesis and analysis of speech.

Comments

University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-91-12.

Surface Structure, Intonation, And Meaning In Spoken Language

MS-CIS-91-12
LINC LAB 193

Mark Steedman

Department of Computer and Information Science
School of Engineering and Applied Science
University of Pennsylvania
Philadelphia, PA 19104-6389

January 1991

Surface Structure, Intonation, and Meaning in Spoken Language*

Mark Steedman

University of Pennsylvania

Abstract

The paper briefly reviews a theory of intonational prosody and its relation syntax, and to certain oppositions of discourse meaning that have variously been called "topic and comment", "theme and rheme", "given and new", or "presupposition and focus." The theory, which is based on Combinatory Categorial Grammar, is presented in full elsewhere. The present paper examines its consequences for the automatic synthesis and analysis of speech.

*Revised September 29, 1992. To appear in M. Bates and R. Weischedel, (eds.), *Challenges in Natural Language Processing*, CUP:Cambridge. The research was supported in part by NSF grant nos. IRI-90-18513 and IRI-91-17110, DARPA grant no. N00014-90-J-1863, and ARO grant no. DAAL03-89-C0031.

The structural units of phrasal intonation are frequently orthogonal to the syntactic constituent boundaries that are recognised by traditional grammar and embodied in most current theories of syntax. As a result, much recent work on the relation of intonation to discourse context and information structure has either eschewed syntax entirely (cf. [7], [15], [22], [8]), or has supplemented traditional syntax with entirely non-syntactic string-related principles (cf. [12]). Recently, Selkirk [54] and others have postulated an autonomous level of "intonational structure" for spoken language, distinct from syntactic structure. Structures at this level are plausibly claimed to be related to discourse-related notions, such as "focus". However, the involvement of two apparently uncoupled levels of structure in Natural Language grammar appears to complicate the path from speech to interpretation unreasonably, and to thereby threaten the feasibility of computational speech recognition and speech synthesis.

In [59] and [60], I argue that the notion of intonational structure formalised by Pierrehumbert, Selkirk, and others, can be subsumed under a rather different notion of syntactic surface structure, that emerges from the "Combinatory Categorial" theory of grammar [57], [58]. This theory engenders surface structure constituents corresponding directly to phonological phrase structure. Moreover, the grammar assigns to these constituents interpretations that directly correspond to what is here called "information structure" – that is, the aspects of discourse-meaning that have variously been termed "topic" and "comment", "theme" and "rheme", "given" and "new" information, and/or "presupposition" and "focus".

The consequent simplification of the path from speech to higher level modules including syntax, semantics, and discourse pragmatics, seems likely to facilitate a number of applications in spoken language understanding. On the analysis side, it can be expected to facilitate the use of such high level modules to "filter" the ambiguities that unavoidably arise from low-level word recognition. On the synthesis side, it can be expected to similarly facilitate the production of intonation contours that are more appropriate to discourse context than the default intonations characteristic of current "text-to-speech" packages. The present paper considers these further implications for speech processing.

1 The Combinatory Grammar of Intonation

1.1 The Problem

One quite normal prosody (b, below) for an answer to the question (a) intuitively imposes the intonational structure indicated by the brackets (stress, marked in this case by raised pitch, is indicated by capitals):

(1) a. I know that Alice likes velvet. But what does MAry prefer?
b. (MA-ry prefers) (CORduroy).

Such a grouping is orthogonal to the traditional syntactic structure of the sentence.

This phenomenon is a property of grammar, and should not be confused with the disruptions caused by hesitations and other performance disfluencies. Intonational structure remains strongly constrained by meaning. For example, contours imposing bracketings like the following are not allowed:

(2) #(Three cats)(in ten prefer corduroy)

Halliday [23] observed that this constraint, which Selkirk [54] has called the "Sense Unit Condition", seems to follow from the *function* of phrasal intonation, which is to convey what will here be called "information structure" – that is, distinctions of focus, presupposition, and propositional attitude towards entities in the discourse model. These discourse entities are more diverse than mere nounphrase or propositional referents, but they do not include such non-concepts as "in ten prefer corduroy."

Among the categories that they *do* include are what Wilson and Sperber and E. Prince [50] have termed "open propositions". One way of introducing an open proposition into the discourse context is by asking a Wh-question. For example, the question in (1), *What does Mary prefer?* introduces an open proposition. As Jackendoff [32] pointed out, it is natural to think of this open proposition as a functional *abstraction*, and to express it as follows, using the notation of the λ-calculus:

(3) $\lambda x\ [(prefer'\ x)\ mary']$

(Primes indicate semantic interpretations whose detailed nature is of no direct concern here.) When this function or concept is supplied with an argument *corduroy'*, it *reduces* to give a proposition, with the same function argument relations as the canonical sentence:

(4) (*prefer' corduroy'*) *mary'*

It is the presence of the above open proposition rather than some other that makes the intonation contour in (1)b felicitous. (That is not to say that its presence uniquely *determines* this response, nor that its explicit mention is necessary for interpreting the response.)

These observations have led linguists such as Selkirk to postulate a level of "intonational structure", independent of syntactic structure and related to information structure. The involvement of two apparently uncoupled levels of structure in natural language grammar appears to complicate the path from speech to interpretation unreasonably, and to thereby threaten a number of computational applications in speech recognition and speech synthesis.

It is therefore interesting to observe that all natural languages include syntactic constructions whose semantics is also reminiscent of functional abstraction. The most obvious and tractable class are Wh-constructions themselves, in which some of the same fragments that can be delineated by a single intonation contour appear as the residue of the subordinate clause. Another and much more problematic class of fragments results from coordinate constructions. It is striking that the residues of wh-movement and conjunction reduction are also subject to something like a "sense unit condition". For example, strings like "in ten prefer corduroy" are as resistant to coordination as they are to being intonational phrases.[1]

(5) *Three cats in twenty like velvet, and in ten prefer corduroy.

Since coordinate constructions constitute another major source of complexity for theories of natural language grammar, and also offer serious obstacles to computational applications, the earlier papers suggest that this conspiracy

[1]I do not claim that such coordinations are absolutely excluded, just that if they are allowed at all then: a) extremely strong and unusual contexts are required, and b) that such contexts will tend to support (2) as well.

between syntax and prosody should be interpreted as evidence for a unified notion of structure that is somewhat different from traditional surface constituency, based on Combinatory Grammar.

1.2 Combinatory Grammars.

Combinatory Categorial Grammar (CCG, [57]) is an extension of Categorial Grammar (CG). Elements like verbs are associated with a syntactic "category" which identifies them as *functions*, and specifies the type and directionality of their arguments and the type of their result. We use a notation in which a rightward-combining functor over a domain β into a range α are written α/β, while the corresponding leftward-combining functor is written $\alpha\backslash\beta$. α and β may themselves be function categories. For example, a transitive verb is a function from (object) NPs into predicates – that is, into functions from (subject) NPs into S:

(6) $prefers := (S\backslash NP)/NP : prefer'$

Such categories can be regarded as encoding the semantic type of their translation, which in the notation used here is identified by the expression to the right of the colon. Such functions can combine with arguments of the appropriate type and position by functional application:

(7)
```
Mary    prefers    corduroy
----   ---------  --------
 NP    (S\NP)/NP      NP
       ---------------->
             S\NP
  -------------<
        S
```

The syntactic types are identical to semantic types, apart from the addition of directional information. The derivation can therefore also be regarded as building a compositional interpretation, ($prefer'$ $corduroy'$) $mary'$, and of course such a "pure" categorial grammar is context free.

Coordination might be included in CG via the following rule, allowing constituents of like type to conjoin to yield a single constituent of the same type:

(8) $X \quad conj \quad X \Rightarrow X$

(9)
```
I    loath   and   detest   velvet
-- --------- ---- --------- ------
NP (S\NP)/NP conj (S\NP)/NP   NP
   ----------------------&
            (S\NP)/NP
```

(The rest of the derivation is omitted, being the same as in (7).) In order to allow coordination of contiguous strings that do not constitute constituents, CCG generalises the grammar to allow certain operations on functions related to Curry's combinators [14]. For example, functions may nondeterministically *compose*, as well as apply, under the following rule:

(10) *Forward Composition:* (>B)
$X/Y : F \quad Y/Z : G \Rightarrow X/Z : \lambda x \; F(Gx)$

The most important single property of combinatory rules like this is that they have an invariant semantics. This one composes the interpretations of the functions that it applies to, as is apparent from the right hand side of the rule.[2] Thus sentences like *I suggested, and would prefer, corduroy* can be accepted, via the following composition of two verbs (indexed as **B**, following Curry's nomenclature) to yield a composite of the same category as a transitive verb. Crucially, composition also yields the appropriate interpretation for the composite verb *would prefer*:

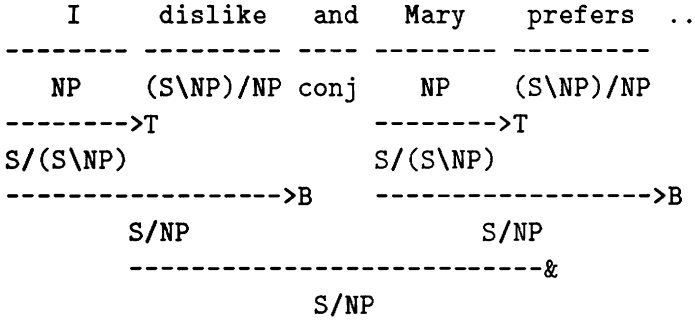
(11)
```
... suggested  and   would   prefer ...
    --------- ---- --------- ------
    (S\NP)/NP conj (S\NP)/VP VP/NP
                   ---------------->B
                     (S\NP)/NP
    -------------------------&
          (S\NP)/NP
```

[2]The rule uses the notation of the λ-calculus in the semantics, for clarity. This should not obscure the fact that it is functional composition itself that is the primitive, not the λ operator.

Combinatory grammars also include type-raising rules, which turn arguments into functions over functions-over-such-arguments. These rules allow arguments to compose, and thereby take part in coordinations like *I dislike, and Mary prefers, corduroy.* They too have an invariant compositional semantics which ensures that the result has an appropriate interpretation. For example, the following rule allows the conjuncts to form as below (again, the remainder of the derivation is omitted):

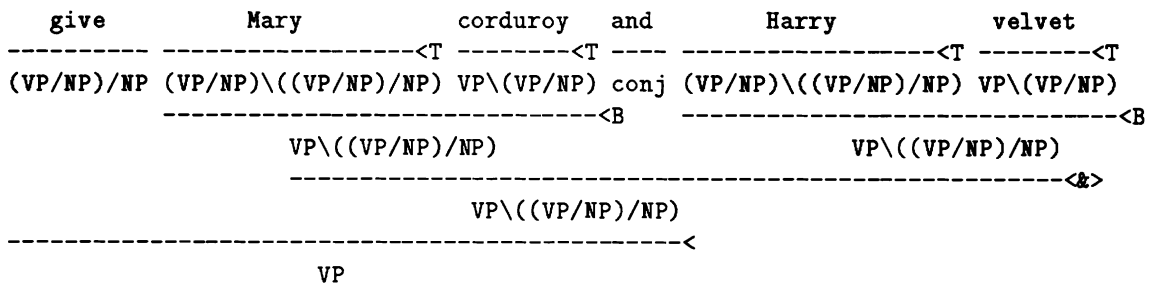
(12) *Subject Type-raising:* (>T)
$NP : y \;\Rightarrow\; S/(S\backslash NP) : \lambda F\, Fy$

(13)



This apparatus has been applied to a wide variety of coordination phenomena, including "left node raising" [18], "backward gapping" in Germanic languages, including verb-raising constructions [56], and gapping, [58]. For example, the following analysis is proposed by Dowty [18] for the first of these:

(14)



The important feature of this analysis is that it uses "backward" rules of type-raising <T and composition <B that are the exact mirror-image of the

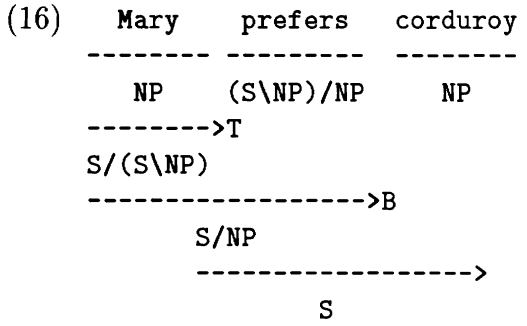
two "forward" versions introduced as examples (10) and (12). It is therefore a prediction of the theory that such a construction can exist in English, and its inclusion in the grammar requires no additional mechanism whatsoever. The earlier papers show that no *other* non-constituent coordinations of dative-accusative NP sequences are allowed in any language with the English verb categories, given the assumptions of CCG. Thus the following are ruled out in principle, rather than by stipulation:

(15) a. *Harry velvet and give Mary corduroy
b. *give corduroy Mary and velvet Harry

A number of related well-known cross-linguistic generalisations concerning the dependency of so-called "gapping" upon lexical word-order are also captured (see Dowty [18] and others [56], [58]).

1.3 Intonation, Parsing, and Context

Examples like the above show that combinatory grammars embody a view of surface structure according to which strings like *Mary prefers* are constituents. It follows, according to this view, that they must also be possible constituents of non-coordinate sentences like *Mary prefers corduroy*, as in the following derivation:

(16)


An entirely unconstrained combinatory grammar would in fact allow any bracketing on a sentence, although the grammars we actually write for configurational languages like English are heavily constrained by local conditions. (An example might be a condition on the composition rule that is

tacitly assumed below, forbidding the variable Y in the composition rule to be instantiated as NP, thus excluding constituents like $*[ate\ the]_{VP/N}$). It nevertheless follows that, for each semantically distinct analysis of a sentence, the involvement of the combinatory operation of functional composition engenders an equivalence class of derivations, which impose different constituent structures but are guaranteed to yield identical interpretations. In more complex sentences than the above, there will be many semantically equivalent derivations for each distinct interpretation.

Such additional non-determinism in grammar, over and above the non-determinism that is usually recognised, creates obvious problems for the parser, and has on occasion been referred to as "spurious" ambiguity. This term is very misleading. Whether or not the present theory is correct, the non-determinism is *there*, in the competence grammar of coordinate constructions, and any parser that actually covers this range of constructions will have to deal with it. It is only the comparitive neglect of these constructions by the parsing community that has led them to ignore this perfectly genuine source of nondeterminism. The papers [45], [59], [65] and [66] discuss the complexity of this problem in the worst case. However, in [13] it is suggested that the evaluation of partial, incomplete, interpretations with respect to a discourse model including a representation of discourse information plays a crucial role. These possibilities will be explored further below.

However the parsing problem is resolved, the interest of such non-standard structures for present purposes should be obvious. The claim is simply that the non-standard surface structures that are induced by the combinatory grammar to explain coordination in English subsume the intonational structures that are postulated by Pierrehumbert *et al.* to explain the possible intonation contours for sentences of English. The claim is that that in spoken utterances, intonation helps to determine *which* of the many possible bracketings permitted by the combinatory syntax of English is intended, and that the interpretations of the constituents that arise from these derivations, far from being "spurious", are related to distinctions of discourse focus among the concepts and open propositions that the speaker has in mind.

The proof of this claim lies in showing that the rules of combinatory grammar can be made sensitive to intonation contour, which limit their application in spoken discourse. We must also show that the major constituents

of intonated utterances like (1)b, under the analyses that are permitted by any given intonation, correspond to the information structure of the context to which the intonation is appropriate, as in (a) in the example (1) with which the proposal begins. This demonstration will be quite simple, once we have established the following notation for intonation contours.

We will use a notation which is based on the theory of Pierrehumbert [46], as modified in more recent work by Selkirk [54], Beckman and Pierrehumbert [6], [47], and Pierrehumbert and Hirschberg [48], and as explicated in the chapter by Pierrehumbert in the present volume. The theory proposed below is in fact compatible with any of the standard descriptive accounts of phrasal intonation. However, a crucial feature of Pierrehumbert's theory for present purposes is that it distinguishes two subcomponents of the prosodic phrase, the *pitch accent* and the *boundary*.[3] The first of these tones or tone-sequences coincides with the perceived major stress or stresses of the prosodic phrase, while the second marks the righthand boundary of the phrase. These two components are essentially invariant, and all other parts of the intonational tune are interpolated. Pierrehumbert's theory thus captures in a very natural way the intuition that the same tune can be spread over longer or shorter strings, in order to mark the corresponding constituents for the particular distinction of focus and propositional attitude that the melody denotes. It will help the exposition to augment Pierrehumbert's notation with explicit prosodic phrase boundaries, using brackets. These do not change her theory in any way: all the information is implicit in the original notation.

Consider for example the prosody of the sentence *Mary prefers corduroy* in the following pair of discourse settings, which are adapted from Jackendoff [32, pp. 260]:

(17) Q: Well, what about the CORduroy? Who prefers THAT?

A: (MARy) (prefers CORduroy).
H* L L+H* LH%

[3]For the purposes of this chapter, the distinction between the intonational phrase proper, and what Pierrehumbert and her colleagues call the "intermediate" phrase, will be largely suppressed. However, these categories differ in respect of boundary tone-sequences – see the chapter by Pierrehumbert in the present volume – and the distinction is implicit below.

(18) Q: Well, what about MARy? What does SHE prefer?

A: (MARy prefers) (CORduroy).
L+H* LH% H* LL%

In these contexts, the main stressed syllables on both *Mary* and *corduroy* receive a pitch accent, but a different one. In the former example, (17), there is a prosodic phrase on *Mary* made up of the pitch accent which Pierrehumbert calls H*, immediately followed by an L boundary. There is another prosodic phrase having the pitch accent called L+H* on *corduroy*, preceded by null or interpolated tone on the words *prefers*, and immediately followed by a boundary which is written LH%. (I base these annotations on Pierrehumbert and Hirschberg's [48, ex. 33] discussion of a similar example.)[4] In the second example (18) above, the two tunes are reversed: this time the tune with pitch accent L+H* and boundary LH% is spread across a prosodic phrase *Mary prefers*, while the other tune with pitch accent H* and boundary LL% is carried by the prosodic phrase *corduroy* (again starting with an interpolated or null tone).[5]

The meaning that these tunes convey is intuitively very obvious. As Pierrehumbert and Hirschberg point out, the latter tune seems to be used to mark some or all of that part of the sentence expressing information that the speaker believes to be *novel to the hearer.* In traditional terms, it marks the "comment" – more precisely, what Halliday called the "rheme". In contrast, the L+H* LH% tune seems to be used to mark some or all of that part of the sentence which expresses information which in traditional terms is the "topic" – in Halliday's terms, the "theme".[6] For present purposes, a theme can be thought of as conveying *what the speaker assumes to be the subject of mutual interest*, and this particular tune marks a theme as *novel to the conversation as a whole*, and as standing in a contrastive relation to the previous theme. (If the theme is not novel in this sense, it receives *no* tone

[4]We continue for the moment to gloss over Pierrehumbert's distinction between "intermediate" and "intonational" phrases.

[5]The reason for notating the latter boundary as LL%, rather than L reflects the distinction between intonational and intermediate phrases.

[6]The concepts of theme and rheme are distantly related to Grosz et al's [21] concepts of "backward looking center" and "forward looking center".

in Pierrehumbert's terms, and may even be left out altogether.)[7] Thus in (18), the L+H* LH% phrase including this accent is spread across the phrase *Mary prefers.*[8] Similarly, in (17), the same tune is confined to the object of the open proposition *prefers corduroy*, because the intonation of the original question indicates that prefering corduroy *as opposed to some other stuff* is the new topic or theme.[9]

The L+H* LH% intonational melody in example (18) belongs to a phrase *Mary prefers ...* which corresponds under the combinatory theory of grammar to a grammatical constituent, complete with a translation equivalent to the open proposition $\lambda x[(prefer'\ x)\ mary']$. The combinatory theory thus offers a way to derive such intonational phrases, using only the independently motivated rules of combinatory grammar, entirely under the control of appropriate intonation contours like L+H* LH%.

1.4 Combinatory Prosody

The L+H* LH% intonational melody in example (18) belongs to a phrase *Mary prefers ...* which corresponds under the combinatory theory of grammar to a grammatical constituent, complete with a translation equivalent to the open proposition $\lambda x[(prefer'\ x)\ mary']$. The combinatory theory thus offers a way to derive such intonational phrases, using only the independently motivated rules of combinatory grammar, entirely under the control of appropriate intonation contours like L+H* LH%.[10]

One extremely simple way to do this is the following. We interpret the two pitch accents as functions over boundaries, of the following types:

[7]Here I depart slightly from Halliday's definition. The present proposal also follows Lyons [38] in rejecting Hallidays' claim that the theme must necessarily be sentence-initial.

[8]An alternative prosody, in which the contrastive tune is confined to *Mary*, seems equally coherent, and may be the one intended by Jackendoff. I believe that this alternative is informationally distinct, and arises from an ambiguity as to whether the topic of this discourse is *Mary* or *What Mary prefers.* It too is accepted by the rules below.

[9]Note that the position of the pitch accent in the phrase has to do with a further dimension of information structure within both theme and rheme, which it is tempting to call "focus" but safer to call "emphasis". I ignore this dimension here.

[10]This section is a simplified summary of the fuller accounts presented in [59] and [60].

(19) L+H* := *Theme/Bh*
H* := *rheme/bl*
H* := *Rheme/Bl*

– that is, as functions over boundary tones into the two major informational types, the Hallidean "Theme" and "Rheme". The Rheme is further distinguished as *Rheme* or *rheme*, according to the type of its boundary, a distinction which reflects its status as an intonational or intermediate phrase. The reader may wonder at this point why we do not replace the category *Theme* by a functional category, say *Utterance/Rheme*, corresponding to its semantic type. The answer is that we do not want this category to combine with anything but a *complete* rheme. In particular, it must not combine with a function into the category *Rheme* by functional composition. Accordingly we give it a non-functional category, and supply the following special purpose prosodic combinatory rules:[11]

(20) *Theme* *Rheme* ⇒ *Utterance*
rheme *Theme* ⇒ *Utterance*

We next define the various boundary tones as arguments to these functions, as follows:

(21) LH% := *Bh*
LL% := *Bl*
L := *bl*

Finally, we accomplish the effect of interpolation of other parts of the tune by assigning the following polymorphic category to all elements bearing no tone specification, which we will represent as the tone Ø:

(22) Ø := *X/X*

Syntactic combination can then be made subject to the following simple restriction:

[11]This pair of rules is a rather crude simplification for the sake of brevity of the account in [59] and [60].

(23) *The Prosodic Constituent Condition:* Combination of two syntactic categories via a syntactic combinatory rule is only allowed if their prosodic categories can *also* combine.

(The prosodic and syntactic combinatory rules need not be the same).

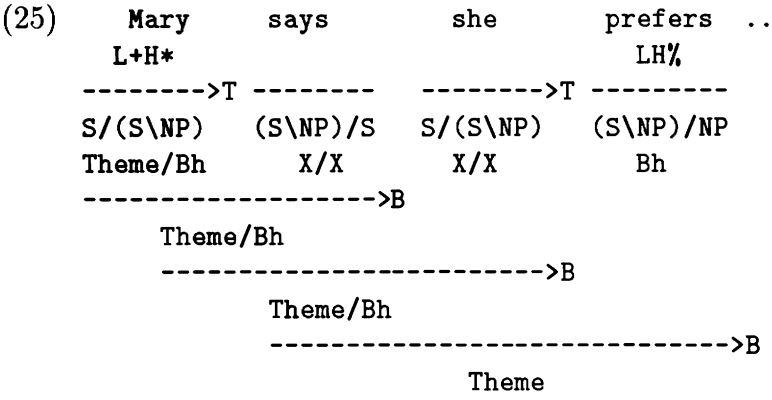
This principle has the sole effect of excluding certain derivations for spoken utterances that would be allowed for the equivalent written sentences. For example, consider the derivations that it permits for example (18) above. The rule of forward composition is allowed to apply to the words *Mary* and *ate*, because the prosodic categories can combine (by functional application):

(24)

```
      Mary                 prefers        ...
     L+H*                    LH%
-------------          -----------------
  NP:mary'             (S\NP)/NP:prefer'
  Theme/Bh                   Bh
------------------>T
S/(S\NP): \P[P mary']
  Theme/Bh
-------------------------------------->B
       S/NP: \X[(prefer' X) mary']
              Theme
```

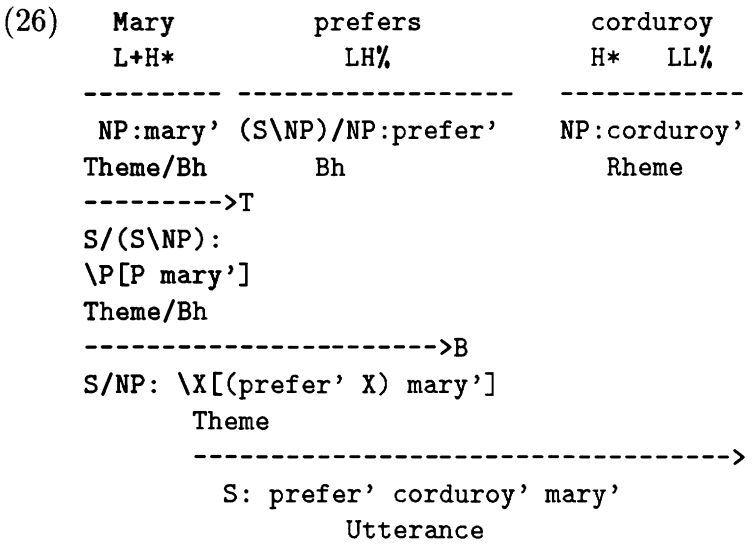
The category X/X of the null tone allows intonational phrasal tunes like L+H* LH% tune to spread across any sequence that forms a grammatical constituent according to the combinatory grammar. For example, if the reply to the same question *What does Mary prefer?* is *MARY says she prefers CORduroy*, then the tune will typically be spread over *Mary says she prefers* ... as in the following (incomplete) derivation, in which much of the syntactic and semantic detail has been omitted in the interests of brevity:

(25)



The rest of the derivation of (18) is completed as follows, using the first rule in ex. (20):

(26)

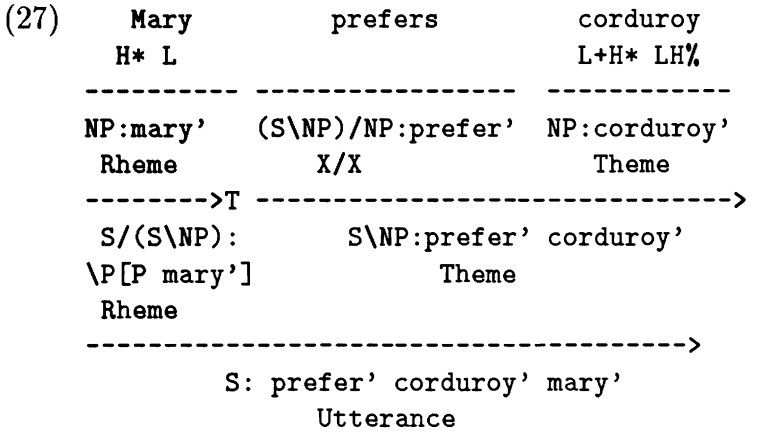


The division of the utterance into an open proposition constituting the theme and an argument constituting the rheme is appropriate to the context established in (18). Moreover, the theory permits no *other* derivation for this intonation contour. Of course, repeated application of the composition rule, as in (25), would allow the L+H* LH% contour to spread further, as in *(MARY says she prefers)(CORduroy.*

In contrast, the parallel derivation is forbidden by the prosodic constituent condition for the alternative intonation contour on (17). Instead,

the following derivation, excluded for the previous example, is now allowed:

(27)



No other analysis is allowed for (27). Again, the derivation divides the sentence into new and given information consistent with the context given in the example. The effect of the derivation is to annotate the entire predicate as an L+H* LH%. It is emphasised that this does *not* mean that the *tone* is spread, but that the whole constituent is marked for the corresponding discourse function — roughly, as contrastive given, or theme. The finer grain information that it is the object that is contrasted, while the verb is given, resides in the tree itself. Similarly, the fact that boundary sequences are associated with words at the lowest level of the derivation does not mean that they are *part of* the word, or specified in the lexicon, nor that the word is the entity that they are a boundary *of*. It is prosodic phrases that they bound, and these also are defined by the tree.

All the other possibilities for combining these two contours in a simple sentence are shown elsewhere [59] to yield similarly unique and contextually appropriate interpretations.

Sentences like the above, including marked theme and rheme expressed as two distinct intonational/intermediate phrases are by that token unambiguous as to their information structure. However, sentences like the following, which in Pierrehumbert's' terms bear a single intonational phrase, are much more ambiguous as to the division that they convey between theme and rheme:

(28) (I read a book about CORduroy)
H* LL%

Such a sentence is notoriously ambiguous as to the open proposition it presupposes, for it seems equally appropriate as a response to any of the following questions:

(29) a. What did you read a book about?
b. What did you read?
c. What did you do?

Such questions could in suitably contrastive contexts give rise to themes marked by the L+H* LH% tune, bracketing the sentence as follows:

(30) a. (I read a book about)(CORduroy)
b. (I read)(a book about CORduroy)
c. (I)(read a book about CORduroy)

It seems that we shall miss a generalisation concerning the relation of intonation to discourse information unless we extend Pierrehumbert's theory very slightly, to allow prosodic constituents resembling *null* intermediate phrases, without pitch accents, expressing unmarked themes. Since the boundaries of such intermediate phrases are not explicitly marked, we shall immediately allow all of the above analyses for (28). Such a modification to the theory can be introduced by the following rule, which nondeterministically allows constituents bearing the null tone to become a theme:

(31) $X/X \Rightarrow$ *Theme*

The rule is nondeterministic, so it correctly continues to allow a further analysis of the entire sentence as a single Intonational Phrase conveying the Rheme. Such an utterance is the appropriate response to yet another open-proposition establishing question, *What happened?*.)

The following observation is worth noting at this point, with repect to the parsing problem for CCG (see section 2.1.2) above. The above rule introduces nondeterminism into the intonational grammar, just when it looked as though intonation acted to eliminate non-determinism from the syntax.

However, the null tone is used precisely when the theme is entirely mutually known, and established in the context. It follows that the this nondeterminism *only arises when the hearer can be assumed to be able to resolve it on the basis of discourse context.* This observation is in line with the results of [3], which suggest that the resolution of non-determinism by reference to discourse context is an important factor in human parsing for both written and spoken language, a matter to which we return in the second part of the paper.

With the generalisation implicit in the above rule, we are now in a position to make the following claim:

(32) The structures demanded by the theory of intonation and its relation to contextual information are the same as the surface syntactic structures permitted by the combinatory grammar.

Because constructions like relativisation and coordination are more limited in the derivations they require, often forcing composition, rather than permitting it, a number of corollaries follow, such as the following:

(33) Anything which can coordinate can be an intonational constituent, and *vice versa.*

and

(34) Anthing which can be the residue of relativisation can be an intonational constituent.

These claims are discussed further in [59].

2 Applications to Speech Processing

Under the present theory, the pathway between the speech-wave and the sort of logical form that can be used to interrogate a database is as in Figure 1. Such an architecture is considerably simpler than the one that is implicit in the standard theories. Phonological form now maps via the rules of

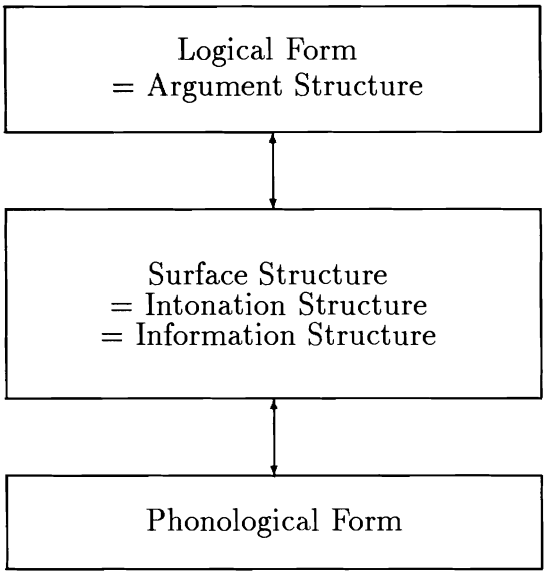


Figure 1: Architecture of a CCG-based Prosody

combinatory grammar directly onto a surface structure, whose highest level constituents correspond to intonational constituents, annotated as to their discourse function. Surface structure is therefore isomorphic to intonational structure. It also subsumes information structure, since the translations of those surface constituents correspond to the entities and open propositions which constitute the topic or theme (if any) and the comment or rheme. These in turn reduce via functional application to yield canonical function-argument structure, or "logical form".[12] There are a number of obvious potential advantages for the automatic synthesis and recognition of spoken language in such a theory, and perhaps it is not to early to speculate a little on how they might be realised.

[12]This term is used loosely. We have said nothing here about how questions of quantifier scope are to be handled, and we assume that they are derived from this representation at a deeper level still.

2.1 Intonation and the Analysis of Spoken Language

The most important potential application for the theory lies in the area of speech recognition. Where in the past parsing and phonological processing have tended to deliver conflicting phrase-structural analyses, and have had to be pursued independently, they now are seen to be in concert. The theory therefore offers the possibility that simply structured modular processors which use both sources of information at once will one day be more easily devised. That is not of course to say that intonational cues remove all local structural ambiguity. Nor is it to underestimate the other huge problems that must be solved before this potential can be realised. But such an architecture may reasonably be expected to simplify the problem of resolving local structural ambiguity in both domains, for the following reason.

First, why is practical speech recognition hard? There seem to be two reasons. One is that the discrete segmental or word-level representations that provide the input to processes of comprehension are realised in the speech wave as the result of a highly non-linear physical system in the form of the vocal tract and its muscular control. This system has many of the computational characteristics of a "relaxation" process of the kind discussed by (for example) Hinton [27], in which a number of autonomous but interacting parallel motor processes combine by an interative approximating procedure to achieve a cooperative result. (In Hinton's paper, this kind of algorithm is used to control reaching by a jointed robot). In the speech domain, this sort of system, in which the articulators act in concert to produce the segments, the result is the phenomenon of "coarticulation", which causes the realisation of any given ideal segment to depend upon the neighbouring segments in very complex ways. It is very hard to invert the process, and to work backwards from the resulting speechwave to the underlying abstract segments that are relevant to higher levels of analysis.

For this reason, the problem of automatically recognising intonational cues such as pitch accents and boundary tones should not be underestimated. The acoustic realisation in the fundamental frequency F_0 of the intonational tunes discussed above is entirely dependent upon the rest of the phonology – that is, upon the phonemes and words that bear the tune. In particular: the realisation of boundary tones and pitch accents is heavily dependent on segmental effects, so that the former can be confounded with the latter.

Moreover F_0 itself may be locally undefined, due to non-linearities and chaotic effects in the vocal tract.[13] (For example, the realisation of the tune H* LL% on the two words "TitiCAca" and "CineRAma" is dramatically different.) It therefore seems most unlikely that intonational contour can be identified in isolation from word recognition. The converse also applies: intonation contour effects the acoustic realisation of words, particularly with respect to timing. It is therefore likely that the benefits of combining intonational recognition and word recognition will eventually be mutual, and will extend the benefits that already accrue to stochastic techniques for word recognition (cf. [33], [35], [36]). As Pierrehumbert has pointed out, part of their success stems from the way in which Hidden Markov Models represent a combination of prosodic and segmental information.

However, such techniques alone may well not be enough to support practical general purpose speech recognition, because of a second source of difficulty in speech recognition. Acoustic information seems to be exceedingly underspecified with respect to the segments. As a result, the output of phonetic- or word- recognition processes is genuinely ambiguous, and characterised by numerous acoustically plausible but spurious alternative candidates. This is probably not just an artifact of the current speech recognition algorithms. It is very likely that the best we shall be able to do with low level analysis alone on the waveform corresponding to a phrase like "recognise speech", even taking account of coarticulation with intonation, will be to produce a table of candidates that might be orthographically represented as follows. (The example is made up, and is adapted from Henry Thompson. But I think it is a fair representation):

(35) wreck# a# nice# beach
recognise # speech
wreck# on# ice# beach
wreck# an# eyes# peach
recondite's # beach
recondite # speech
reckon# nice# speech
...

[13] While smoothing algorithms go some way towards mitigating the latter effects, they are not completely effective.

– and these are only the candidates that constitute lexical words.

Such massive ambiguity is likely to completely swamp higher level processing unless it can be rapidly eliminated. It seems likely that the way that this is done is by "filtering" the low level candidates on the grounds of coherence at higher levels of analysis, such as syntactic and semantic levels. This is the mechanism of "weak" or selective interaction between modules proposed in [13], [3], according to which the higher level is confined to sending "interrupts" to lower level processes, causing them to be abandoned or suspended, but cannot otherwise affect the autonomy of the lower level. They and Fodor [20] contrast such models with the "strong" interaction, which compromises modularity by allowing higher levels to direct the inner workings of the lower, affecting the actual analyses that get proposed in the first place.

Thus one might expect that syntactic well-formedness could be used to select among the word candidates, in much the same way that we assumed above that the lexicon would be used to reject incoherent strings of phonemes. However, inspection of the example suggests that syntax alone may not be much help, for all of the above word strings are syntactically coherent. (The example is artificial, but it is typical in this respect). It is only at the level of semantics that many of them can be ruled out, and only at the level of pragmatics that in a context like the present discussion all but one can be excluded as incoherent.

However, nondeterminism at low levels of analysis must be eliminated quickly, or it will swamp the processor at that level. It follows that we would like to begin this filtering process as early as possible, and therefore need to "cascade" processors at the different levels, so that the filtering process can begin while the analysis is still in progress. Since we have noted that syntax alone is not going to do much for us, we need semantics and pragmatics to kick in at an early stage, too. The resultant architecture can be viewed as in Figure 2.

Since the late 'seventies, in work by such as Carroll et al. [9], Marslen-Wilson et al. [41], Tanenhaus [62], and Swinney [61]), a increasing number of studies have shown that some such architecture is in fact at work, and in [3] and [13], it is suggested that the weak interaction bears the major responsibility for resolving nondeterminism in syntactic processing. However, for such a mechanism to work, all levels must be *monotonically* related – that

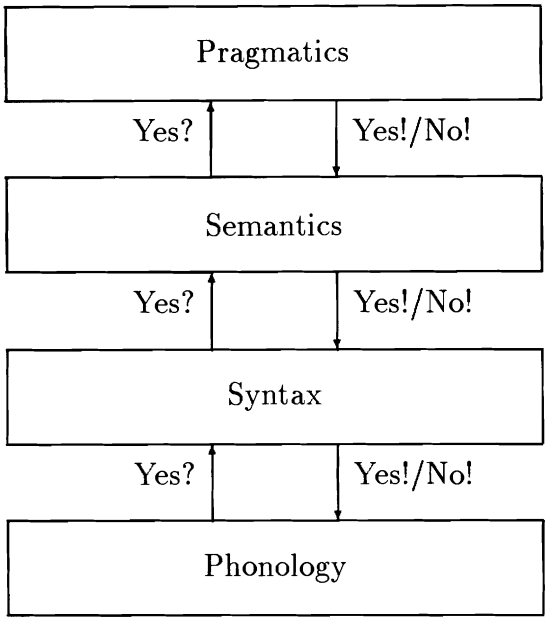


Figure 2: Architecture of a Weakly Interactive Processor

is, rules must be essentially declarative and unordered, if partial information at a low level is to be useable at a higher level.

The present theory has all of the requisite properties. Not only is syntactic structure closely related to the structure of the speech signal, and therefore easier to use to "filter" the ambiguities arising from lexical recognition. More importantly, the constituents that arise under this analysis are also semantically interpreted. These interpretations have been shown above to be directly related to the concepts, referents and themes that have been established in the context of discourse, say as the result of a question. These discourse entities are in turn directly reducible to the structures involved in knowledge-representation and inference. The direct path from speech to these higher levels of analysis offered by the present theory should therefore make it possible to use more effectively the much more powerful resources of semantics and domain-specific knowledge, including knowledge of the discourse, to filter low-level ambiguities, using larger grammars of a

more expressive class than is currently possible. While vast improvements in purely bottom-up word recognition can be expected to continue, such filtering is likely to remain crucial to successful speech processing by machine, and appears to be characteristic of all levels of human processing, for both spoken and written language.

However, to realise the potential of the present theory for the domain of analysis requires a considerable further amount of basic research into significant extensions of available techniques at many levels other than syntax, including the phonological level and the level of Knowledge Representation, related to pragmatics. It will be a long project.

2.2 Discourse-model Driven Synthesis of Intonation

A more immediate return can be expected from the present theory in the form of significant improvements in both acceptability and intelligibility over the fixed or default intonation contours that are assigned by text-to-speech programs like MITalk and its commercial offspring [2]. One of the main shortcomings of current text-to-speech synthesis programs is their inability to vary intonation contour dependent upon context. While considerable ingenuity has been devoted to minimising the undesirable effects, via algorithms with some degree of sensitivity to syntax, and the generation of general-purpose default intonations, this shortcoming is really an inevitable concomitant of the text-to-speech task itself. In fact, a truly general solution to the problem of assigning intonation to unconstrained text is nothing less than a solution to the entire problem of understanding written Natural Language. We therefore propose the more circumscribed goal of generating intonation from a known discourse model in a constrained and well-understood domain, such as inventory management, or travel planning.[14]

[14]The proposal to drive intonation from context or the model is of course not a new one. Work in the area includes an early study by Young and Fallside, [67], and more recent studies by Houghton, Isard and Pearson (cf. [28], [29], [30], [31]), and by Davis and Hirschberg (cf. [17]) on synthesis of intonation in context, and by Yoshimara Sagisaka [53], although the representations of information structure and its relation to syntax that these authors use are quite different from those we propose. The work of t'Hart et al. at IPO ([25], [26], [63]) and that implicit in the MITalk algorithm itself ([44], [2]) do not make explicit reference to information structure, and are more indirectly relevant.

The inability to vary intonation appropriately affects more than the mere æsthetic qualities of synthetic speech. On occasion, it affects intelligibility as well. Consider the following example, from an inventory management task

EXAMPLE: The context is as follows: *A storekeeper carries a number of items including Widgets and Wodgets. The storekeeper and his customer are aware that Widgets and Wodgets are two different kinds of advanced pencil-sharpener, and that the 286 and 386 processors are both suitable for use in such devices. The latter is of course a faster processor, but it will transpire that the customer is unaware of this fact.* The following conversation ensues:[15]

(36) Q1: Do you carry PENCIL-sharpeners?
L* LH%

A1: We carry WIDgets, and WODgets.
H* H H* LL%

For storekeepers to be asked and to answer questions about the stock that they carry is expected by both parties, so both utterances have an unmarked theme λX *carry'* X *storekeeper'*, signalled by null tone on the relevant substring. The question includes a marked rheme, concerning pencil sharpeners. The response also includes a marked rheme, concerning specific varieties of this device. The dialogue continues:

[15]Once again, we use Pierrehumbert's notation to make the tune explicit. However, the contours we have in mind should be obvious from the context alone and the use of capitals to indicate stress.

(37) Q2: Which pencil-sharpener has a THREE-eight-six PROcessor?
H* H* LH% H* H* LL%

A2: WODGets have a THREE-eight-six PROcessor
H* L L+H* L+H* LH%

Q3: WHAT PROcessor do WIDgets have?
H* H* LH% H* LL%

A3: WIDGets have a TWO-eight-six processor.
L+H* LH% H* LL%

The two responses A2 and A3 are almost identical, as far as lexical items and traditional surface structure go. However, the context has changed in between, and the intonation should change accordingly, if the sentence is to be easily understood. In the first case, answer A2, the theme, which might be written $\lambda X[(have'386')X]$, has been established by the previous *Wh*-question Q2. This theme is in contrast to the previous one (which concerned varieties of pencil-sharpeners), and is therefore intonationally marked.[16] (Only a part of the theme was emphasised in Q2, so the same is true in A3). However, the next *Wh*-question Q3 establishes a new theme, roughly, $\lambda X[(have'X)widget']$. Since it is again different to the previous theme, it is again marked with the tune L+H* LH%.[17]

It is important to observe that comprehension would be seriously impeded if the two intonational tunes were exchanged.

The dialogue continues with the following exchange (recall that Wodgets are the device with the faster processor):[18]

[16] An unmarked theme bearing the null tone seems equally appropriate. However, it is as easy (and much safer) for the generator to err on the side of over-specificity.

[17] Again, an unmarked theme with null tone would be a possible (but less cooperative) alternative. However, the position of the pitch-accent would remain unchanged.

[18] The example is adapted to the present domain from a related example discussed by [48].

(38) Q4: Are WODgets FASter than Widgets?
H* H* LH%

A4: The three-eight-SIX machine is ALways faster.
L+H* LH% H* LL%

The expression "the three eight six machine" refers to the Wodget, because of contextually available information. Accordingly, it is marked as such by the L+H* LH% tune, and the predicate is marked as rheme. The answer therefore amounts to a positive answer to the question. It simultaneously conveys the *reason* for the answer. (To expect that a question-answering program for a real database could exhibit such cooperative and conversationally adept responses is not unreasonable – see papers in [34] and [5] – although it may go beyond the capability of the system we shall develop for present purposes.)

Contrast the above continuation with the following, in which a similarly cooperative response is negative:

(39) Q4': Are WIDgets FASter than Wodgets?
H* H* LH%

A4': The three-eight-SIX machine is always FASter
H* L L+H* LH%

The expression *the three eight six machine* refers again to Wodgets, but this time it does *not* correspond to the theme established by Q4'. Accordingly, an H* pitch accent is used to mark it as part of the rheme, *not* part of the theme established by Q4'. Note that A4 and A4' are identical strings, but that exchanging their intonation contours would again result in both cases in infelicity, caused by the failure of the presupposition that Widgets are a three-eight-six - based machine. In this case, any given default intonation, say one having an unmarked theme and final H*LL%, will force one of the two readings, and will therefore mislead the hearer.

How might such a system be brought into being? The analysis of spoken language is, as we have seen, a problem in it own right, to which we briefly return below. But within the present framework one can readily imagine a query system which process either written or spoken language concerning

some simple and widely studied domain, such as the "inventory management" domain illustrated above, the "travel agent" domain that has been studied in a number of recent projects, or the "route-finding" domain used by Davis and Hirschberg [17], to yield analyses of the present kind, related to the information structure of the query. Such domains are quite adequate to motivate the distinctions of information structure that are of interest here, and could readily be extended to include aspects of the "intelligent user-manual" paradigm, as in the last example. Quite modest extensions to incorporate open propositions as individuals in the model would provide opportunities to use intonation contours whose discourse function is the correction of misconceptions, without enlarging the knowledge representation problem unduly.

ANALYSING QUERIES: Such a query system would parse and interpret the questions according to a combinatory grammar, to produce interpretations including a representation of information structure, including distinctions of theme, rheme and focus, associated with interpretations such as open propositions and arguments, as well as a traditional function-argument structure. For example, the parser might deliver something like the following analysis for question Q3 above, *What processor do Widgets have?*:[19]

(40) Function/Argument-Structure = $\lambda X[(processor'X)\&((have'X)widget')]$
Theme = S/(S/NP): $\lambda Pred[\lambda X[(processor'X)\&(Pred\ X)]]$
Rheme = S/NP: $\lambda X[((have'X)widget')]$

Such a representation could be used in two ways. First, it could be used to update a discourse model by establishing the corresponding discourse entities in the model. Second, it could be used to derive an answer to the question, the function-argument structure being used to interrogate a simple relational database of facts to yield an answer, perhaps looking something like the following:

(41) Function/Argument-Structure = $(processor'386')\&((have'386')widget')$

[19]The example is based on the output of a prototype parser written in Prolog using a simplified Montague-style semantics. Interpretations again appear to the right of syntactic categories, separated by a colon. Again the use of the lambda calculus is a notational convenience – the system itself uses a different unification-based representation for categories and their interpretations, following [58], and uses combinators as applicative primitives.

The discourse representation and this answer to the database query could then be used to generate entirely from scratch a representation of a response, including a representation of its information structure, the latter including all distinctions of theme and focus that are relevant to specifying its intonation contour, as follows.

GENERATING RESPONSES: It seems reasonable to assume initially (no doubt oversimplifying with respect to real human generators of utterances) that the discourse representation and the query between them deterministically specify the response, and that no backtracking or replanning of the utterance of the complex kinds discussed by [39] will be involved. In particular, it seems reasonable initially to assume that *the Rheme of the original question determines the Theme of the answer*, so that some structure such as the following can be used as the input to a generator:

(42) utterance(theme(S/NP: $\lambda X[((have'X)widget')]$), rheme(NP: 386′))

This structure will then be used to determine by rule a complete specification of the phonological form of the corresponding string, including all details of pitch and timing, in a form suitable for input to the speech synthesiser itself.

The question of whether entities like *widget* and *386processor* should be expressed in the form of NPs like "Widgets" and "the 386 processor", or as pronouns, or as more complex NPs, is of course also determined by discourse context. The much fuller discourse representations envisaged in the present system could also be exploited to make these finer "tactical" decisions as well ([64], [39], [16]). Promising candidates for attention in this regard are cleft constructions, ellipses, and the coordinate constructions, all of which provided the original motivation for combinatory grammars (see section 1.2 above), and all of which are strongly constrained by discourse information and by intonation. They would be required for examples like the following, in the inventory management domain:

(43) Q: Do Widgets have a 386 processor?
A: It is Wodgets that have a 386 processor.

(44) Q: Do both pencil sharpeners include a serial port?
A: Widgets do, and Wodgets do not, include an RS232 interface.

(45) Q: What processor do Widgets and Wodgets have?
A: Widgets have a 286 processor, and Wodgets, a 386 processor.

A further promising area for investigation lies in the interaction of intonation with "focussing operators" like *only* and *even*, and with semantic notions of scope, as evinced in examples like the following (cf. [52], [55]):

(46) Q: Do all pencil-sharpeners have a serial port?
A: Only Widgets have a serial port.

The rules for specifying phonological form, including pitch and timing, remain to be specified within the CCG framework, and are a subject for further research. One set of techniques that could be used in at least a preliminary application, and which fall short of full synthesis-by-rule, are to be found in the literature of Concatenative text-to-speech Synthesis using LPC-based, and other, techniques (cf. [43], [19], [40], [51], [1], [24], [10]).

References

[1] Allen, Jonathan: "Short-term spectral analysis, synthesis, and modification by discrete Fourier transform," *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, 25, pp.235-238.

[2] Allen, Jonathan, Sharon Hunnicutt, and Dennis Klatt: 1987, *From Text to Speech: the MITalk system*, Cambridge, University Press.

[3] Altmann, Gerry and Mark Steedman: 1988, 'Interaction with Context During Human Sentence Processing' *Cognition*, 30, 191-238.

[4] Anderson, M, J. Pierrehumbert, M. Liberman: 1984, 'Synthesis by rule of English intonation patterns', *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, 1984.

[5] Appelt, Doug: 1985, "Planning English Referring Expressions' *Artificial Intelligence*, 26, 1-33.

[6] Beckman, Mary and Janet Pierrehumbert: 1986, 'Intonational Structure in Japanese and English', *Phonology Yearbook*, 3, 255-310.

[7] Bolinger, Dwight: 1972, 'Accent is Predictable (If You're a Mind Reader)', *Language*, 48, 633-644.

[8] Brown, Gillian, and George Yule: 1983, *Discourse Analysis*, Cambridge, University Press.

[9] Carroll, J. and Tom Bever: 1978, 'The Perception of Relations', in William J. M. Levelt and Giovanni Flores d'Arcais (eds.), *Studies in the Perception of Language*, Wiley, New York NY.

[10] Charpentier, F. and E. Moulines, 'Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones,' Proceedings EUROSPEECH89, vol. 2, pp. 13-19.

[11] Chomsky, Noam: 1970, 'Deep Structure, Surface Structure, and Semantic Interpretation', in D. Steinberg and L. Jakobovits, *Semantics*, CUP, Cambridge, 1971, 183-216.

[12] Cooper, William and Julia Paccia-Cooper: 1980, *Syntax and Speech*, Harvard University Press, Cambridge MA.

[13] Crain, Stephen and Mark Steedman: 1985, 'On not being led up the garden path: the use of context by the psychological parser', in D. Dowty, L. Kartunnen, and A. Zwicky, (eds.), *Natural Language Parsing: Psychological, Computational and Theoretical Perspectives*, ACL Studies in Natural Language Processing, Cambridge University Press, 320-358.

[14] Curry, Haskell and Robert Feys: 1958, *Combinatory Logic*, North Holland, Amsterdam.

[15] Cutler, Anne, and Stephen Isard: 1980, 'The Production of Prosody', in Brian Butterworth, (ed.), *Language Production, Vol. 1*, New York, Wiley, 246-269.

[16] Dale, Robert: 1989, 'Cooking up Referring Expressions', *Proceedings of the 27th Annual Conference of the ACL*, Vancouver, 68-75. June 1989.

[17] Davis, James and Julia Hirschberg: 1988, 'Assigning Intonational Features in Synthesised Spoken Directions', *Proceedings of the 26th Annual Conference of the ACL*, Buffalo, 187-193.

[18] Dowty, David: 1988, Type raising, functional composition, and non-constituent coordination, in Richard T. Oehrle, E. Bach and D. Wheeler, (eds), *Categorial Grammars and Natural Language Structures*, Reidel, Dordrecht, 153–198.

[19] Fallside, Frank and S.J. Young: 1978, 'Speech Output from a Computer-controlled Water-supply Network', *Proc. IEEE*, 125, 157-161.

[20] Fodor, Gerry: 1983, *The Modularity of Mind*, MIT Press, Cambridge MA.

[21] Grosz, Barbara, Aravind Joshi, and Scott Weinstein: 1983, 'Providing a Unified Account of Definite Noun Phrases in Discourse, *Proceedings of the 21st Annual Conference of the ACL*, Cambridge MA, July 1983, 44-50.

[22] Gussenhoven, Carlos: 1983, *On the Grammar and Semantics of Sentence Accent*, Dordrecht, Foris.

[23] Halliday, Michael: 1967, *Intonation and Grammar in British English*, Mouton, The Hague.

[24] Hamon, C. et al: 1989 'A diphone synthesis system based on time-domain prosodic modifications of speech,' ICASSP89, 238-241.

[25] 't Hart, J. and A. Cohen: 1973, 'Intonation by Rule: a Perceptual Quest', *Journal of Phonetics*, 1, 309-327.

[26] 't Hart, J. and R. Collier: 1975, 'Integrating Different Levels of Phonetic Analysis, *Journal of Phonetics*, 3, 235-255.

[27] Hinton, Geoffrey: 1984, 'Parallel Computation for Controlling an Arm', *Journal of Motor Behaviour*, 16, 171-194.

[28] Houghton, George: 1986, *The Production of Language in Dialogue: a Computational Model*, unpublished Ph.D dissertation, University of Sussex.

[29] Houghton, George and Stephen Isard: 1987 'Why to speak, what to say, and how to say it', in P. Morris (ed.), *Modelling Cognition*, Wiley. (1987)

[30] Houghton, George and M. Pearson: 1988, 'The Production of Spoken Dialogue,' in M. Zock and G. Sabah (eds), *Advances in Natural Language Generation: An Interdisciplinary Perspective, Vol. 1*, Pinter Publishers, London.

[31] Stephen Isard and M. Pearson: 1988 'A repertoire of British English intonation contours for synthetic speech', *Proceedings of Speech '88, 7th FASE Symposium, Edinburgh, 1988*, pp.1233-1240.

[32] Jackendoff, Ray: 1972, *Semantic Interpretation in Generative Grammar*, MIT Press, Cambridge MA.

[33] Jelinek, Fred: 1976, 'Continuous Speech Recognition by Continuous Methods', *Proceedings of Institute of Electrical and Electronic Engineers*, 64, 532-556.

[34] Joshi, Aravind, Bonnie Lynn Webber, and Ivan Sag (Eds.): 1981, *Elements of Discourse Understanding*, Cambridge, University Press.

[35] Lee, Kai-Fu: 1989, *Automatic Speech Recognition*, Kluwer, Dordrecht.

[36] Lee, Kai-Fu: 1990, 'Context-dependent Phonetic Hidden Markov Models for Continuous Speech Recognition', *IEEE Transactions on Acoustics Speech and Signal Processing*, to appear.

[37] Liberman, Mark and J. Pierrehumbert: 1984, 'Intonational Invariance under Changes in Pitch Range and Length', in M. Aranoff and R. Oehrle, (eds.), *Language Sound Structure: Studies in Phonology Presented to Morris Hallé*, MIT Press, Cambridge MA.

[38] Lyons, John: 1977. *Semantics, vol. II*, Cambridge University Press.

[39] McDonald, David: 1983, 'Description-directed Control', *Computers and Mathematics*, 9, 111-130.

[40] Markel, John, and Augustine Gray: 1976, *Linear Prediction of Speech*, Springer-Verlag, Berlin.

[41] Marslen-Wilson, William, Lorraine K. Tyler and Mark Seidenberg: 1978, 'The Semantic Control of Sentence Segmentation', in William J. M. Levelt and Giovanni Flores d'Arcais (eds.), *Studies in the Perception of Language*, Wiley, New York NY.

[42] Moens, Marc, and M. Steedman: 1987, 'Temporal Ontology and Temporal Reference', *Journal of Computational Linguistics*, 14, 15–28.

[43] Olive, J.P. and L. Nakatani: 1974, 'Rule Synthesis by Word-concatenation: a First Step', *Journal of the Acoustical Society of America*, 55, 660-666.

[44] O'Shaughnessy, D.: 1977, 'Fundamental Frequency by Rule for a Text-to-Speech System', *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, IEEE Cat. No. 77CH1197-3 ASSP, New York, IEEE, 568-570.

[45] Pareschi, Remo, and Mark Steedman. 1987. A lazy way to chart parse with categorial grammars, *Proceedings of the 25th Annual Conference of the ACL*, Stanford, July 1987, 81–88.

[46] Pierrehumbert, Janet: 1980, *The Phonology and Phonetics of English Intonation*, Ph.D dissertation, MIT. (Dist. by Indiana University Linguistics Club, Bloomington, IN.)

[47] Pierrehumbert, Janet, and Mary Beckman: 1989, *Japanese Tone Structure*, MIT Press, Cambridge MA.

[48] Pierrehumbert, Janet, and Julia Hirschberg, 1987, 'The Meaning of Intonational Contours in the Interpretation of Discourse', ms. Bell Labs.

[49] Pereira, Fernando and Martha Pollack: 1990, 'Incremental Interpretation', ms., AT&T Bell Labs, Murray Hill NJ/SRI International, Menlo Park CA.

[50] Prince, Ellen F. 1986. On the syntactic marking of presupposed open propositions. Papers from the Parasession on Pragmatics and Grammatical Theory at the 22nd Regional Meeting of the Chicago Linguistic Society, 208-222.

[51] Rabiner, L. and R. Schafer: 1976, 'Digital Techniques for Computer Voice Response: Implementations and Applications', *Proceedings of Institute of Electrical and Electronic Engineers*, 64, 416-433.

[52] Rooth, Mats: 1985, *Association with Focus*, unpublished PhD dissertation, University of Massachusetts, Amherst.

[53] Sagisaka, Yoshinori : 1990, 'On the Prediction of Global F0 Shape for Japanese Text-to-Speech,' ICASSP 90, pp. 325-328.

[54] Selkirk, Elisabeth: 1984, *Phonology and Syntax*, MIT Press, Cambridge MA.

[55] von Stechow, Arnim: 1989, 'Focussing and Backgrounding Operators', Fachgruppe Sprachwissenschaft der Universität Konstanz, Arbeitspapier Nr. 6.

[56] Steedman, Mark: 1985a. Dependency and Coordination in the Grammar of Dutch and English, *Language* 61.523-568.

[57] Steedman, Mark: 1987. Combinatory grammars and parasitic gaps. *Natural Language & Linguistic Theory*, 5, 403-439.

[58] Steedman, Mark: 1990. 'Gapping as Constituent Coordination', *Linguistics & Philosophy*, 13, 207-263.

[59] Steedman, Mark: 1991a, 'Structure and Intonation', *Language*, 68, 260-296.

[60] Steedman, Mark: 1991b, 'Syntax, Intonation and Focus', in E. Klein and F. Veltmann, (eds.) *Natural Language and Speech: Proceedings of the Symposium, ESPRIT Conference, Brussels, Nov. 1991*, Springer Verlag, Berlin. 21-38.

[61] Swinney, David: 1979. Lexical Access during Sentence Comprehension: (re)consideration of context effects. *Journal of Verbal Learning and Verbal Behaviour*, 18, 645-660.

[62] Tanenhaus, Michael: 1978, *Sentence Context and Sentence Perception*, Ph D thesis, Columbia University.

[63] Terken, Jacques: 1984, 'The distribution of accents in instructions as a function of discourse structure', *Language and Speech*, 27, 269-289.

[64] Thompson, Henry: 1977, 'Strategy and Tactics in Language Production', *Proceedings of the 13th Annual Conference of the Chicago Linguistics Society*, Chicago IL, April, 1977.

[65] Vijay-Shankar, K and David Weir: 1990, 'Polynomial Time Parsing of Combinatory Categorial Grammars', *Proceedings of the 28th Annual Conference of the ACL*, Pittsburgh, June 1990.

[66] Wittenburg, Kent: 1987, 'Predictive Combinators: a Method for Efficient Processing of Combinatory Grammars', *Proceedings of the 25th Annual Conference of the ACL*, Stanford, July 1987, 73–80.

[67] Young, S. and F. Fallside: 1979,'Speech Synthesis from Concept: a Method for Speech Output from Information Systems' *Journal of the Acoustical Society of America*, 66, 685-695.